Big Data and Data Science in Critical Care

L. Nelson Sanchez-Pinto, MD; Yuan Luo, PhD; and Matthew M. Churpek, MD, PhD

The digitalization of the health-care system has resulted in a deluge of clinical big data and has prompted the rapid growth of data science in medicine. Data science, which is the field of study dedicated to the principled extraction of knowledge from complex data, is particularly relevant in the critical care setting. The availability of large amounts of data in the ICU, the need for better evidence-based care, and the complexity of critical illness makes the use of data science techniques and data-driven research particularly appealing to intensivists. Despite the increasing number of studies and publications in the field, thus far there have been few examples of data science projects that have resulted in successful implementations of data-driven systems in the ICU. However, given the expected growth in the field, intensivists should be familiar with the opportunities and challenges of big data and data science. The present article reviews the definitions, types of algorithms, applications, challenges, and future of big data and data science in critical care. CHEST 2018; 154(5):1239-1248

KEY WORDS: big data; critical care; data science; machine learning; prediction models

The digitalization of the health-care system is changing the way we practice medicine and conduct clinical research.[1,2] The widespread implementation of electronic health records (EHRs) is paving the way for big data research and is bringing the world of data science to the patient's bedside.[2-4] Within the health-care system, the ICU presents a particularly convincing case for using data science to improve patient care.[5] The evidence supporting many of the interventions performed in the ICU is scarce, and practice variability is abundant.[5,6] In addition, the complexity of critical illness makes the traditional reductionist approach to medical research insufficient; that is, single-drug intervention trials or single pathway biomarker studies are unlikely to satisfy the clinical realities of the ICU.[5] Critical care research requires an integrative approach that embraces the complexity of critical illness and the computational technology and algorithms that can make it possible.[7,8] Moreover, the data required to follow this approach are being generated and digitized in troves. EHRs, bedside monitors, medication pumps, and ventilators are continuously generating new minable data, and soon the advancement of modern molecular diagnostics will result in a deluge of "omics" data derived from the genome, transcriptome, microbiome, and a long list of other "-omes" (Fig 1).

As big data and data science gradually infiltrate most aspects of clinical research and, ultimately, clinical care in the ICU, it is

ABBREVIATIONS: AUC = area under the receiver-operating characteristic curve; EHR = electronic health record; MIMIC = Multiparameter Intelligent Monitoring in Intensive Care; NLP = natural language processing

AFFILIATIONS: From the Department of Pediatrics (Critical Care) (Dr Sanchez-Pinto) and the Department of Preventive Medicine (Health and Biomedical Informatics) (Drs Luo and Sanchez-Pinto), Northwestern University Feinberg School of Medicine, Chicago, IL; and the Department of Medicine (Dr Churpek), The University of Chicago, Chicago, IL.

CORRESPONDENCE TO: Matthew M. Churpek, MD, PhD, Pulmonary and Critical Care, University of Chicago, 5841 S Maryland Ave, MC 6023, Chicago, IL 60637; e-mail: matthew.churpek@uchospitals.edu

DOI: https://doi.org/10.1016/j.chest.2018.04.037

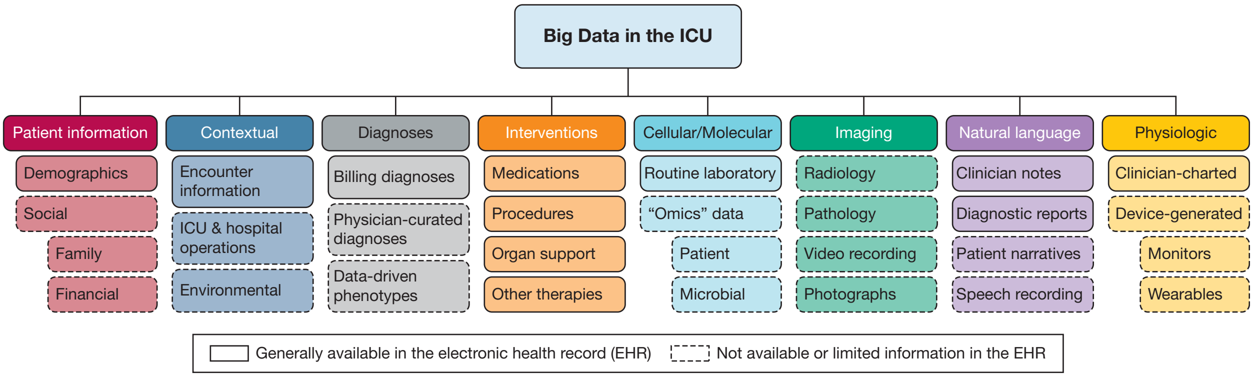

Figure 1 – *Some of the major sources of big data in the ICU. The term "omics" refers to the data derived from modern molecular techniques (eg, genomics, transcriptomics, proteomics, metabolomics, microbiomics). EHR = electronic health record.*

increasingly evident that intensivists should be familiar with the promise and perils of these approaches. The present article reviews the definitions, types of algorithms, applications, challenges, and future of big data and data science in critical care.

Definitions in Data Science

Big data can be defined as digital data that are generated in high volume and high variety and that accumulate at high velocity, resulting in datasets too large for traditional data-processing systems.[2,9] In practice, big data in health care depend on both the breadth and the depth of the data being captured. For example, administrative health-care datasets with few data elements per patient record (low depth) are usually considered big data problems when they contain millions of records (wide breadth). Conversely, when applying next-generation sequencing and other "-omics" approaches (high depth), just a few dozen patients can become a big data problem (narrow breadth).[10]

Data science can be defined as "the set of fundamental principles that support and guide the principled extraction of information and knowledge from data."[9] A closely related term is data mining, which is the actual extraction of knowledge from data via machine learning algorithms that incorporate data science principles. Machine learning is the field of study that focuses on how computers learn from data and the development of algorithms that make this learning possible.[11] Finally, another important concept in data science is domain expertise, which in health care can be defined as the understanding of real-world clinical problems and the realities of patient care that help frame and contextualize the application of data science to health-care problems.[8,9,11]

These and other pertinent definitions in data science are presented in Table 1.

The Ecosystem of Data Science in Health Care

The data revolution in health care would not be possible if it were not for several key developments, including: (1) the data science movement that has transformed other industries; (2) the extraordinary growth in computational power; (3) the availability of open source tools and low-cost equipment to perform advanced analyses; and (4) the increasing availability of educational resources and advanced degrees in data science and related fields.[12] Open source programming and scripting languages, such as R and Python, have extensive libraries of statistical packages and machine learning algorithms that are relatively easy to use by researchers with some training and have greatly democratized the access to data science techniques.

Educational resources, such as graduate programs in data science, are also increasingly available. Notable among these are massive open online courses, which have led to the increased popularity of data science and the applications of machine learning techniques to real-world problems.[13] There are now a multitude of courses from prestigious institutions that cover data science and machine learning, and many of them are free. These courses provide the interested researcher with educational opportunities from world-class experts at the touch of a button. Other websites contain user-created code to run machine learning algorithms from scratch (eg, https://github.com/) or host data science competitions for participants around the world (eg, https://www.kaggle.com/). This rich online environment is ideal for data scientists to learn and grow, with the

TABLE 1] Definitions of Common Terms in Data Science

Term	Definition
Big data	Digital data that are generated in high volume and high variety and that accumulate at high velocity, resulting in datasets too large for traditional data-processing systems
Data science	The set of fundamental principles that support and guide the principled extraction of information and knowledge from data
Data mining	The extraction of knowledge from data via machine learning algorithms that incorporate data science principles
Domain expertise	The understanding of real-world problems in a given domain (eg, critical care medicine) that helps frame and contextualize the application of data science to solve these problems
Machine learning	The field of study that focuses on how computers learn from data and the development of algorithms that make this learning possible
Features	The data elements, also known as independent variables, used to train a model. Features can be simple transformations of the raw data (eg, average heart rate in the last 24 h) or complex transformation such as the ones performed by neural networks (see Table 2)
Outcomes	The data elements, also known as dependent variables, represent the target for training in a supervised learning model. Outcomes can be categorical (eg, yes/no) or continuous (eg, length of hospital stay). Categorical binary outcomes are the most common in medicine (eg, died or alive by 28 days). Binary outcomes are typically represented as a Boolean logic (ie, true/false or 1/0) but can also be represented using fuzzy logic (ie, a range of probabilities, or degrees of truth, between 0 and 1)
Supervised learning	Algorithms that are used to uncover the relationship between a set of features and one or more known outcomes
Unsupervised learning	Algorithms that are used to uncover naturally occurring patterns or groupings in the data, without targeting a specific outcome
Model training	The process through which machine learning algorithms develop a model of the data by learning the relationships between features and, in supervised learning, between features and outcomes. This is also referred to as model derivation or data fitting
Model validation	The process of measuring how well a model fits new, independent data. For example, evaluating the performance of a supervised model at predicting an outcome in new data. This approach is also referred to as model testing.
Predictive model	A model generally trained to predict the likelihood of a condition, event, or response. The US Food and Drug Administration specifically considers predictive strategies as those geared toward identifying groups of patients more likely to respond to an intervention
Prognostic model	A model specifically trained to predict the likelihood of a condition-related endpoint or outcome such as mortality. In general, the goal is to estimate a prognosis given a set of baseline features, regardless of what ultimately leads to the outcome
Overfitting	The phenomenon that occurs when an algorithm learns from idiosyncrasies in the training data, usually referred to as noise. Noisy data are data that are randomly present in the training dataset but do not represent the generalizable truth (usually referred to as signal) that explains the relationships between the features and the outcomes. Overfitting will generally lead to poor performance of the model in an independent validation dataset
Digitization	The conversion of something analog or physical (eg, paper documents, printed images) into a digital format (ie, bits or 1s and 0s)
Digitalization	The wide adoption of digital technologies by an organization to leverage their digitized data with the goal of improving operations and performance. The adoption of electronic health records and other digital technologies (eg, picture archiving and communication systems for medical images, pharmacy management systems, billing systems) are examples of digitalization in health care
Data curation	The process of integrating data from different sources, structuring it, authenticating it, and annotating it to ensure its quality, add value, and facilitate its use and reuse
Structured data	Data (usually discrete or numeric) that are easy to search, summarize, sort, and quantify. Examples include vital signs (eg, heart rate) or laboratory test results (eg, CBC)
Unstructured data	Data that do not conform to a prespecified structure, such as a written narrative, images, video, or audio. Unstructured data are generally harder to search, sort, and quantify. Examples include clinician notes, pathology slides, and radiology images

field of medicine benefiting from advances shared around the world freely through the Internet.

Types of Algorithms in Data Science

Machine learning algorithms are generally divided into two categories: supervised and unsupervised.[11] Semi-supervised algorithms represent a hybrid of the two but have been used less often in health-care problems. Finally, deep learning algorithms defy this classification, even though they derive from artificial neural network algorithms, which are generally classified as supervised algorithms. The most defining characteristic of deep learning is their focus on learning data representations (or features) that can then be used in supervised, unsupervised, or semi-supervised problems. The following discussions review these types of algorithms in more detail (as well as in Fig 2 and Table 2).

Supervised Learning Algorithms

Supervised learning algorithms are used to uncover the relationship between variables of interest and one or more target outcomes.[11,14] For supervised problems, the target outcome(s) must be known. For example, if researchers want to know whether a set of clinical features (eg, vital signs, laboratory tests) can predict ICU mortality, they could apply a supervised learning algorithm to a dataset in which each patient record contains the set of clinical features of interest and a label specifying their outcome ("survived" or "not survived" in this case) (Table 1). Examples of supervised learning algorithms include regression-based methods (eg, linear and logistic regression, lasso, elastic net), tree-based methods (eg, classification and regression trees, random forest, gradient boosted trees), k-nearest neighbor, artificial neural networks, and support vector machines (Table 2).[15]

Unsupervised Learning Algorithms

Unsupervised learning algorithms are used to uncover naturally occurring patterns or groupings in the data, without targeting a specific outcome.[11] The most compelling use case of unsupervised learning in health care is in precision medicine, in which the goal is to uncover subsets of patients who share similar clinical or molecular characteristics and are, in theory, more likely to respond to targeted therapies directed at their shared underlying pathobiology.[16,17] For example, an unsupervised learning algorithm may be used to uncover subgroups of patients with sepsis who have distinct molecular and clinical characteristics and will respond differently to specific drugs, such as corticosteroids.[18] Some examples of unsupervised learning algorithms include clustering algorithms (eg, hierarchical clustering, k-means clustering), latent class analysis, and principal component analysis (Table 2).[11,14,19]

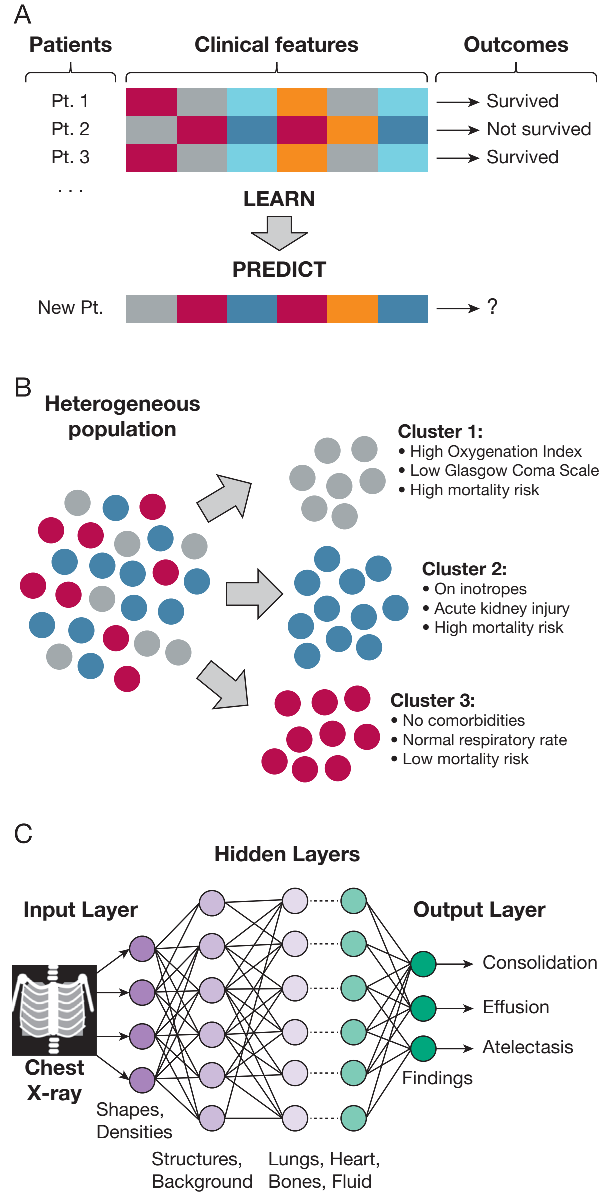

Figure 2 – *A-C, Types of machine learning algorithms applicable to critical care. A, Supervised learning algorithms can be used, for example, to uncover the relationship between patient clinical features (eg, laboratory tests and vital signs) and mortality to predict the outcome in future cases. B, Unsupervised learning algorithms can be used to uncover naturally occurring groupings or clusters of patients based on their clinical characteristics, without targeting a specific outcome. C, Deep learning algorithms can be used, for example, to extract meaningful features from imaging data (eg, chest radiograph) to represent information in an increasingly higher order of hierarchical complexity and be able to make predictions, such as the presence of pathologic findings.*

Deep Learning Algorithms

Deep learning algorithms are designed to extract meaningful features from the data to represent

TABLE 2] Examples of Algorithms Use in Data Science

Algorithm Class	Examples	Description
Classic regression	Linear regression, logistic regression	Linear regression is a supervised learning algorithm that models the relationship between one or more features and a continuous outcome by fitting a regression line that minimizes the sum of all the residuals, which are the distances between each feature in the training data and the line being fitted to model them. Logistic regression is a generalization of the linear model that uses the logistic function to estimate the probability of a binary outcome. To do this, the fitted sigmoid-shaped curve of the logistic function maps the feature values into a probability between 0 and 1
Regularized regression	Lasso, ridge regression, elastic net	An extension of the classic regression algorithms in which a penalty is imposed to the fitted model to reduce its complexity and decrease the risk of overfitting (see Table 1).
Tree-based	Classification and regression trees, random forest, gradient boosted trees	A class of supervised learning algorithm based on decision trees. Decision trees are a sequence of "if-then-else" splits that are derived by iteratively separating the data into groups based on the relationship of the features with the outcome. Random forest and gradient boosted trees are example of ensemble tree models. Ensemble models combine the output of many trained models to estimate an outcome
Support vector machines	Linear, polynomial, radial basis kernel	A class of supervised learning algorithms that represents the data in a multidimensional feature space and then fits a "hyperplane" that best separates the data based on the outcomes of interest
K-nearest neighbor	K-nearest neighbor	A type of supervised learning algorithm that represents data in multidimensional feature space and uses local information about observations closest to a new example to predict the outcome for that example
Bayesian	Naive Bayes, Bayesian network	A class of supervised learning algorithms that use Bayes' theorem of conditional probability, which is the probability that something will happen given that something else has already occurred. In general, Bayesian algorithms work by iteratively updating the probability of an outcome (or posterior belief) given new data
Neural network	Artificial neural network, deep neural network	A class of nonlinear algorithms built using layers of nodes that extract features from the data and perform combinations that best represent the underlying structure, usually to predict an outcome. Neural networks can be shallow (eg, a perceptron with two layers) or deep (multiple layers), which form the basis for the field of deep learning
Dimensionality reduction algorithms	Principal component analysis, linear discriminant analysis	A class of unsupervised learning algorithms that exploit the inherent structure in the data to describe data using less information. Principal components, for example, summarize a large set of correlated features into a smaller number of representative features
Latent class analysis	Latent class analysis	A type of unsupervised learning algorithm that identifies unseen subgroups, or latent classes, in the data. Class membership is unknown for each example so the probability of class membership is indirectly estimated by measuring the patterns in the data
Cluster analysis	K-means, hierarchical cluster analysis	A class of unsupervised learning algorithm that uses the inherent structures in the data to best organize the data into subgroups of maximum commonality based on some distance measure between features

information in an increasingly higher order of hierarchical complexity in the form of stacked layers of nodes (or "neurons").[20] For example, if the input is a photo of several people, the first layer of nodes might simply extract straight lines, curves, and color hues. Deeper layers may combine some of those lines, curves, and hues to represent eyes, noses, ears, and other more complex features. After deeper layers of nodes have perceived increasingly more complex features in an unsupervised way, they can then be used to perform specific tasks, such as matching the faces in the photo to certain specific people with known features. In medical applications, deep learning has been used, for example, to detect diabetic retinopathy on funduscopic images,[21] detect cancer in skin photographs,[22] or predict clinical outcomes by using EHR data.[23,24]

Data Science Applications in Critical Care

Predictive and Prognostic Models

The most common applications of data science to critical care problems are predictive and prognostic models using supervised learning algorithms. Although identical from a modeling perspective, predictive and prognostic models can be distinguished semantically by the fact that predictive models are generally trained to predict the likelihood of a condition, event, or response, whereas prognostic models are specifically trained to predict the likelihood of a condition-related endpoint or outcome, such as mortality (Table 1).[16,25] This distinction, however, is not always clear in the literature and, depending on the use case, might be irrelevant.

One of the oldest and best-known prognostic models to estimate risk of mortality in ICU patients is the Acute Physiology and Chronic Health Evaluation score, which was first developed in the 1980s by Knaus and colleagues[26,27] using logistic regression. Since then, many other groups have developed predictive and prognostic models using larger, more granular datasets and applying modern machine learning methods. For example, Churpek and colleagues[28] developed a logistic regression model in a dataset of > 250,000 hospital admissions that accurately estimated the risk for ICU transfer, cardiac arrest, or death in ward patients. In a follow-up study, the same group showed that more modern machine learning methods, such as random forests and gradient boosted machines, could more accurately predict clinical deterioration compared with classic logistic regression.[15] In another example, Joshi and Szolovits[29] used 54 clinical variable time series to predict 30-day mortality in ICU patients in the publicly available Multiparameter Intelligent Monitoring in Intensive Care (MIMIC) dataset. They clustered the physiological measurements into organ-specific patient states and achieved a state-of-the-art 30-day mortality prediction area under the receiver-operating characteristic curve (AUC) of 0.91.

Predictive models aimed at identifying patients with specific conditions or those more likely to respond to a specific therapy are more commonly used in the field of oncology, with multiple examples of biomarker-based models used to diagnose particular subtypes of cancer that respond to targeted therapy.[4,16] However, there are some examples in the critical care literature, particularly in sepsis and septic shock.[30,31] For example, Wong and colleagues[18] used a combination of a classification and regression tree-based biomarker risk model and gene expression profiles in pediatric patients with sepsis to identify a subgroup of patients who were more likely to benefit from corticosteroids.

Clustering and Phenotyping

Unsupervised learning algorithms in critical care have mainly been used to uncover naturally occurring subgroups or clusters of patients who share similar clinical and/or molecular characteristics. These clusters are oftentimes called phenotypes, subphenotypes, or subtypes, although there is still little consensus on the terminology.[16] For example, Calfee and colleagues[32] applied latent class analysis and identified two subphenotypes of ARDS using clinical and cytokine data from two randomized controlled trials of ARDS. The subphenotypes identified had distinct differences in inflammatory profiles, response to ventilator strategies, and clinical outcomes. Knox and colleagues[33] used self-organizing maps and k-means clustering to identify four distinct clusters of patients with sepsis-associated multiple organ dysfunction syndrome that were independently associated with outcomes after adjusting for severity of illness. Luo and colleagues[34] analyzed multiple physiological variable trends of patients in the MIMIC dataset and applied nonnegative matrix factorization to group-related trends, which were shown to effectively predict 30-day mortality while maintaining model interpretability. Finally, Vranas and colleagues[35] applied clustering analysis to discover and validate six clinically recognizable subgroups of ICU patients who differed significantly in all baseline characteristics and clinical trajectories despite sharing common diagnoses.

Applications With Nontraditional Data Types

Natural language processing: Much of the data used in critical care studies, such as vital signs or laboratory test results, are structured data that can be easily entered into a relational database or spreadsheet and be sorted and summarized. However, there is a significant amount of clinical information contained in the form of unstructured clinical narratives (eg, progress notes, discharge summaries, nursing notes, diagnostic reports).[36] Methods for analyzing narrative data, generally known as natural language processing (NLP), are designed to extract features from texts that can then be used in task-specific algorithms for different purposes (eg, prognostic modeling). Lehman and colleagues[37] used clinical data and unstructured progress notes from the first 24 h of ICU admissions to estimate the risk of in-hospital mortality. They inferred topic models from progress notes and achieved an AUC of 0.82, which was superior to severity of illness scores based only on structured clinical variables. Ghassemi and colleagues[38] further investigated the prognostic power of topics as features from the first 24 h and achieved an AUC of 0.85 for in-hospital mortality when combining text and structured data. Weissman and colleagues[39] applied NLP to analyze discharge documents of ARDS survivors and found that ARDS itself is rarely mentioned in those documents, as opposed to more frequent mentions of "mechanical ventilation" and "ICU stay." Conversely, their NLP-based document classifier reported 100% accuracy for ARDS identification, suggesting that NLP can be used to effectively identify patients with certain types of conditions.

Physiological waveform analysis: Physiological waveform data from bedside monitors and wearable devices are increasingly being used in data science studies in critical care. Many institutions collect and store physiological monitor data, such as electrocardiography, photoplethysmography, impedance pneumography, invasive arterial manometry, end-tidal capnography, and electroencephalography. The publicly available MIMIC databases contain physiological waveform data for ICU patients at Beth Israel Deaconess Medical Center, which has facilitated the development of the state-of-the-art waveform analysis in the field.[40] For example, researchers have used waveform data to estimate cardiac output data using pulse contour analysis techniques,[41] detect hypovolemia using photoplethysmography data,[42] and predict hyperlactatemia using combined physiologicaldata.[43]

Image analysis: The advancement in the field of deep learning, which is particularly useful for image analysis, has resulted in a rapid increase in the number of studies in this area in the last few years.[44] However, none of the current published studies has tested the usefulness of automated image analysis in an ICU setting. The rapid growth of this field, however, will undoubtedly result in many uses applicable to critical care situations. Perhaps most pertinent to critical care clinicians is the advancement in techniques to detect pulmonary pathology in chest radiographs,[45,46] as well as normal and abnormal findings in brain and abdominal imaging.[44] These techniques could be particularly helpful in ICUs with limited availability of specialists who can accurately interpret radiographic images in a timely fashion, but their effectiveness and safety should first be thoroughly tested before any clinical implementation is considered.

Challenges and Pitfalls of Data Science in Critical Care

Like most emerging technologies, the products of data science research in critical care will undoubtedly go through a series of hype and disillusionment cycles before becoming accepted, proven assets in the study and care of critically ill patients. One of the first challenges that data science faces in critical care is that, despite the increasing number of studies and publications in the field, thus far there have been few examples of data science projects that have resulted in successful implementation of data-driven systems in the ICU.[11] This lack of exposure in the clinical setting inevitably results in a degree of mistrust by clinicians in these data-driven systems.[47,48] Although clinicians are happy to use similar systems to browse their smart televisions, shop online, or interact with social media apps, they are wary of the idea of sharing clinical decision-making responsibilities with machine learning algorithms, particularly if they view them as "black boxes."[48] It is likely that only the implementation of well-designed, interpretable, and effective data-driven systems in the ICU will make clinicians start to gain trust in them. Furthermore, the implementation of these data-driven systems must be performed under the rigorous auspices of well-controlled experimental studies, including (but not limited to) simulation testing, preintervention and postintervention studies, and randomized controlled trials. The medical informatics literature has good examples of using scientifically rigorous approaches to the implementation and testing of digital solutions such as clinical decision support tools and can serve as a model to follow.[49-51]

Clinicians and researchers appraising a data-driven system and the literature that supports it must be aware of common pitfalls that can raise concerns about its value. The effectiveness of a data-driven system goes beyond a measure of performance, such as an AUC or a *P* value. To be effective, a data-driven system must produce actionable outputs for the right patients, at the right time. For example, the output can be predictive information that can help a clinician decide the most effective treatment for a particular patient as soon as a diagnosis is made. Furthermore, when evaluating the clinical implementation of such a system, it is important to know whether it has been tested in an experimental setting and whether it has shown a meaningful impact in a population similar to the one for which it is being considered.

Unfortunately, bad data science abounds. We must make a collective effort to ensure that only good data science evolves into data-driven systems that can be safely tested and used in critically ill patients. The ease of access to large amounts of data and computing power can lead to data mining "fishing expeditions" that can result in low-quality research.[3] Poorly framed clinical problems, bad data, or debatable methods will result in flawed data science, and it can create more problems than it solves.[3,6,48] Using epidemiologic best practices to analyze retrospective data, including thoughtfully adjusting for confounding variables, is just as important in large datasets as it is in smaller ones. In addition, a model may fit well only on the training data but generalize poorly to other data, a phenomenon known as overfitting (Table 1). Overfitting may occur when algorithms learn from idiosyncrasies, or noise, in the training data; techniques such as cross-validation and regularization can be used to mitigate this problem.[52,53]

Poorly implemented digital technologies can harm patients,[54] and only a rigorous approach to their evaluation and implementation can mitigate this risk. Partnerships between data scientists, clinical domain experts, medical informaticians, and implementation science specialists will result in more effective and safer data-driven systems. Clinicians with data science skills, clinical research expertise, and an intimate knowledge of the clinical realities in the ICU can help data science teams capture the right data, address the right clinical problems, and produce the right actionable knowledge.[10] Furthermore, clinician input can help minimize the number of unnecessary alerts or prompts these systems might produce, thereby reducing the risk of alert fatigue, which is another common problem among front-line providers working with novel digital technologies.[55]

Another common concern among clinicians is the perceived loss of autonomy in the face of increasingly more sophisticated computational systems. This concern exists despite the fact that clinicians will readily acknowledge that the complexity of medicine nowadays far exceeds the capacity of the unaided human mind and that perhaps these novel computational systems can help manage some of this complexity.[10] To put it in perspective, humans typically make decisions using fewer than six data points, because anything more than that becomes cognitively too expensive.[56] However, an ICU patient can generate thousands of data points in a single day, and when you add fatigue, interruptions, and the clinicians' own cognitive biases, it is not surprising that many clinical decisions end up being suboptimal.[57] Conversely, computers can sift seamlessly through tens of thousands of data points, they can easily analyze complex nonlinear interactions between variables, they never sleep, and they can multitask effortlessly. However, clinical thinking and medical decision-making are not reproducible by current technologies.[10] The qualitative aspect of clinical decision-making—the so-called "art of medicine"—is impossible to model quantitatively. Furthermore, many factors influencing clinical decision-making, including clinical, societal, and personal factors, are not necessarily reflected in the digital records, and thus any output from a data-driven system will need to be first evaluated, interpreted, and enriched by clinicians before any action is taken. However, to achieve a successful partnership between clinicians and computers, we must first improve the skills of bedside clinicians at interpreting and using the output from these data-driven systems.[48]

Finally, another challenge faced by data science teams in critical care is balancing the need for data openness and reproducibility with the demand for data privacy and security. The open data science movement calls for transparent and reproducible research with seamless data-sharing across institutions. Indeed, a recent study showed an alarming lack of reproducibility in data science studies using the same ICU data, which suggests that algorithms, study procedures, computer code, and even datasets should be openly available to ensure reproducibility.[58] However, this data openness must not result in poor data governance, lack of data security, or loss of confidentiality, all of which are necessary to perform ethical research and maintain public trust.[59]

The Future of Big Data and Data Science in Critical Care

We imagine a future in critical care in which data-driven systems and clinicians work hand-in-hand. Large quantities of clinical, physiologic, and "omics" data are analyzed by computational systems and are served to the bedside clinicians in the form of manageable, interpretable, and actionable knowledge that augment the clinician's decision-making capacity. Predictive models perform diagnostic and therapeutic recommendations, while clinicians contextualize these recommendations and coordinate their implementation. False alerts are kept to a minimum and systems are continuously improved through a collaborative and scientifically rigorous approach.

Data science can be transformative. There is a real opportunity that this scenario will become a reality in the near future, but there is still a lot of work ahead of us. Our patients entrust us with their precious data and we—clinicians, researchers, data scientists, and leaders in critical care—have an obligation to use it in the best possible way.

Acknowledgments

Financial/nonfinancial disclosures: The authors have reported to *CHEST* the following: M. M. C. has a patent pending for risk stratification algorithms for hospitalized patients; he is also supported by a career development award from the National Heart, Lung and Blood Institute and a research project grant program award from the National Institute of General Medical Sciences. None declared (L. N. S.-P., Y. L.).

References

1. Smith M, Saunders R, Stuckhardt L, McGinnis JM. *Best Care at Lower Cost: The Path to Continuously Learning Health Care in America*. Washington, DC: National Academies Press; 2013.
2. Bates DW, Saria S, Ohno-Machado L, Shah A, Escobar G. Big data in health care: using analytics to identify and manage high-risk and high-cost patients. *Health Aff (Millwood)*. 2014;33(7):1123-1131.
3. Badawi O, Brennan T, Celi LA, et al. Making big data useful for health care: a summary of the inaugural mit critical data conference. *JMIR Med Inform*. 2014;2(2):e22.
4. Iwashyna TJ, Liu V. What's so different about big data?. A primer for clinicians trained to think epidemiologically. *Ann Am Thorac Soc*. 2014;11(7):1130-1135.
5. Anthony Celi L, Mark RG, Stone DJ, Montgomery RA. "Big data" in the intensive care unit. Closing the data loop. *Am J Respir Critic Care Med*. 2013;187(11):1157-1160.
6. Ghassemi M, Celi LA, Stone DJ. State of the art review: the data revolution in critical care. *Crit Care*. 2015;19:118.
7. Buchman TG, Billiar TR, Elster E, et al. Precision medicine for critical illness and injury. *Crit Care Med*. 2016;44(9):1635-1638.
8. Johnson AE, Ghassemi MM, Nemati S, Niehaus KE, Clifton DA, Clifford GD. Machine learning and decision support in critical care. *Proceedings IEEE*. 2016;104(2):444-466.
9. Provost F, Fawcett T. Data science and its relationship to big data and data-driven decision making. *Big Data*. 2013;1(1):51-59.
10. Obermeyer Z, Lee T. Lost in thought—the limits of the human mind and the future of medicine. *N Engl J Med*. 2017;377(13):1209.
11. Deo RC. Machine learning in medicine. *Circulation*. 2015;132(20):1920-1930.
12. Murdoch TB, Detsky AS. The inevitable application of big data to health care. *JAMA*. 2013;309(13):1351-1352.
13. Hardin J, Hoerl R, Horton NJ, et al. Data science in statistics curricula: preparing students to "think with data." Am Stat 2015; 2015;69(4):343-353.
14. James G, Witten D, Hastie T, Tibshirani R. *An Introduction to Statistical Learning*, Vol. 112. New York, NY: Springer; 2013.
15. Churpek MM, Yuen TC, Winslow C, Meltzer DO, Kattan MW, Edelson DP. Multicenter comparison of machine learning methods and conventional regression for predicting clinical deterioration on the wards. *Crit Care Med*. 2016;44(2):368-374.
16. Seymour CW, Gomez H, Chang CH, et al. Precision medicine for all? Challenges and opportunities for a precision medicine approach to critical illness. *Crit Care*. 2017;21(1):257.
17. Mayhew MB, Petersen BK, Sales AP, Greene JD, Liu VX, Wasson TS. Flexible, cluster-based analysis of the electronic medical record of sepsis with composite mixture models. *J Biomed Inform*. 2018;78:33-42.
18. Wong HR, Atkinson SJ, Cvijanovich NZ, et al. Combining prognostic and predictive enrichment strategies to identify children with septic shock responsive to corticosteroids. *Crit Care Med*. 2016;44(10):e1000-e1003.
19. Luo Y, Ahmad FS, Shah SJ. Tensor factorization for precision medicine in heart failure with preserved ejection fraction. *J Cardiovasc Transl Res*. 2017;10(3):305-312.
20. Goodfellow I, Bengio Y, Courville A. Deep learning. *Adapt Comput Mach Le*. 2016:1-775.
21. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. 2016;316(22):2402-2410.
22. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542(7639):115-118.
23. Miotto R, Li L, Kidd BA, Dudley JT. Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. *Sci Rep*. 2016;6:26094.
24. Aczon M, Ledbetter D, Ho L, et al. Dynamic mortality risk predictions in pediatric critical care using recurrent neural networks. *arXiv preprint arXiv:170106675*. 2017 Jan 23.
25. Wong HR. Intensive care medicine in 2050: precision medicine. *Intensive Care Med*. 2017;43(10):1507.
26. Knaus WA, Zimmerman JE, Wagner DP, Draper EA, Lawrence DE. APACHE-acute physiology and chronic health evaluation: a physiologically based classification system. *Crit Care Med*. 1981;9(8):591-597.
27. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Crit Care Med*. 1985;13(10):818-829.
28. Churpek MM, Yuen TC, Winslow C, et al. Multicenter development and validation of a risk stratification tool for ward patients. *Am J Respir Crit Care Med*. 2014;190(6):649-655.
29. Joshi R, Szolovits P. Prognostic physiology: modeling patient severity in intensive care units using radial domain folding. Paper presented at: American Medical Informatics Association Annual Symposium Proceedings; November 3-7, 2012; Chicago, IL.
30. Henry KE, Hager DN, Pronovost PJ, Saria S. A targeted real-time early warning score (TREWScore) for septic shock. *Sci Translat Med*. 2015;7(299):299ra122.
31. Nemati S, Holder A, Razmi F, Stanley MD, Clifford GD, Buchman TG. An interpretable machine learning model for accurate prediction of sepsis in the ICU. *Critical Care Medicine*. 2018;46(4):547-553.
32. Calfee CS, Delucchi K, Parsons PE, et al. Subphenotypes in acute respiratory distress syndrome: latent class analysis of data from two randomised controlled trials. *Lancet Respir Med*. 2014;2(8):611-620.
33. Knox DB, Lanspa MJ, Kuttler KG, Brewer SC, Brown SM. Phenotypic clusters within sepsis-associated multiple organ dysfunction syndrome. *Intensive Care Med*. 2015;41(5):814-822.
34. Luo Y, Xin Y, Joshi R, Celi L, Szolovits P. Predicting ICU Mortality Risk by Grouping Temporal Trends from a Multivariate Panel of Physiologic Measurements. Paper presented at: Proceedings of the 30th AAAI Conference on Artificial Intelligence; 2016.
35. Vranas KC, Jopling JK, Sweeney TE, et al. Identifying distinct subgroups of ICU patients: a machine learning approach. *Crit Care Med*. 2017;45(10):1607-1615.
36. Sjoding MW, Liu VX. Can you read me now? Unlocking narrative data with natural language processing. *Ann Am Thorac Soc*. 2016;13(9):1443-1445.
37. Lehman LW, Saeed M, Long W, Lee J, Mark R. Risk stratification of ICU patients using topic models inferred from unstructured progress notes. Paper presented at: AMIA annual symposium proceedings; 2012.
38. Ghassemi M, Naumann T, Doshi-Velez F, et al. Unfolding physiological state: mortality modelling in intensive care units. Paper presented at: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2014.
39. Weissman GE, Harhay MO, Lugo RM, Fuchs BD, Halpern SD, Mikkelsen ME. Natural language processing to assess documentation of features of critical illness in discharge documents of acute respiratory distress syndrome survivors. *Ann Am Thoracic Soc*. 2016;13(9):1538-1545.

40. Saeed M, Villarroel M, Reisner AT, et al. Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC-II): a public-access intensive care unit database. *Crit Care Med*. 2011;39(5):952.

41. Sun JX, Reisner AT, Saeed M, Heldt T, Mark RG. The cardiac output from blood pressure algorithms trial. *Crit Care Med*. 2009;37(1):72.

42. Roederer A, Weimer J, DiMartino J, Gutsche J, Lee I. Robust monitoring of hypovolemia in intensive care patients using photoplethysmogram signals. Paper presented at: Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE; 2015.

43. Dunitz M, Verghese G, Heldt T. Predicting hyperlactatemia in the MIMIC II database. Paper presented at: Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE; 2015.

44. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Medical Image Analysis*. 2017;42:60-88.

45. Gonzalez G, Ash SY, Vegas-Sanchez-Ferrero G, et al. Disease staging and prognosis in smokers using deep learning in chest computed tomography. *Am J Respir Crit Care Med*. 2018;197(2):193-203.

46. Bar Y, Diamant I, Wolf L, Greenspan H. Deep learning with non-medical training used for chest pathology identification. Paper presented at: Proc. SPIE2015.

47. Tangri N, Kent DM. Toward a modern era in clinical prediction: the TRIPOD statement for reporting prediction models. *Am J Kidney Dis*. 2015;65(4):530-533.

48. Verghese A, Shah NH, Harrington RA. What this computer needs is a physician: humanism and artificial intelligence. *JAMA*. 2018;319(1):19-20.

49. Pickering BW, Dong Y, Ahmed A, et al. The implementation of clinician designed, human-centered electronic medical record viewer in the intensive care unit: a pilot step-wedge cluster randomized trial. *Int J Med Inform*. 2015;84(5):299-307.

50. Awdishu L, Coates CR, Lyddane A, et al. The impact of real-time alerting on appropriate prescribing in kidney disease: a cluster randomized controlled trial. *J Am Med Inform Assoc*. 2016;23(3):609-616.

51. Bright TJ, Wong A, Dhurjati R, et al. Effect of clinical decision-support systems: a systematic review. *Ann Intern Med*. 2012;157(1):29-43.

52. Hinton GE, Srivastava N, Krizhevsky A, Sutskever I, Salakhutdinov RR. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:12070580:*2012.

53. Bishop CM. *Pattern recognition and Machine Learning*. New York, NY: Springer Science+Business Media LLC; 2006.

54. Han YY, Carcillo JA, Venkataraman ST, et al. Unexpected increased mortality after implementation of a commercially sold computerized physician order entry system. *Pediatrics*. 2005;116(6):1506-1512.

55. Kizzier-Carnahan V, Artis KA, Mohan V, Gold JA. Frequency of passive EHR alerts in the ICU: another form of alert fatigue [published online ahead of print June 22, 2016]? *J Patient Saf*. https://doi.org/10.1097/PTS.0000000000000270.

56. Kahneman D. *Thinking, Fast and Slow*. New York, NY: Farrar, Staus and Giroux; 2011.

57. Neuraz A, Guerin C, Payet C, et al. Patient mortality is associated with staff resources and workload in the ICU: a multicenter observational study. *Crit Care Med*. 2015;43(8):1587-1594.

58. Johnson AE, Pollard TJ, Mark RG. Reproducibility in critical care: a mortality prediction case study. Paper presented at: Machine Learning for Healthcare Conference; 2017.

59. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. *Health Inf Sci Syst*. 2014;2:3.